



Mimicable embodied demonstration in a decomposed sequence

Two aspects of recipient design in professionals' video-mediated encounters

Due, Brian Lystgaard; Lange, Simon Bierring; Nielsen, Mie Femø; Jarlskov, Celine

Published in:
Journal of Pragmatics

DOI:
[10.1016/j.pragma.2019.07.015](https://doi.org/10.1016/j.pragma.2019.07.015)

Publication date:
2019

Document version
Peer reviewed version

Document license:
[Unspecified](#)

Citation for published version (APA):
Due, B. L., Lange, S. B., Nielsen, M. F., & Jarlskov, C. (2019). Mimicable embodied demonstration in a decomposed sequence: Two aspects of recipient design in professionals' video-mediated encounters. *Journal of Pragmatics*, 152, 13-27. <https://doi.org/10.1016/j.pragma.2019.07.015>

**Mimicable embodied demonstration in a decomposed sequence:
Two aspects of recipient design in professionals' video-mediated encounters**

Brian L. Due, Simon Bierring Lange, Mie Femø Nielsen, Celine Jarlskov

This is a post-print version of the paper published in Journal of Pragmatics 152 (2019) 13-27
<https://doi.org/10.1016/j.pragma.2019.07.015>

Mimicable embodied demonstration in a decomposed sequence: Two aspects of recipient design in professionals' video-mediated encounters

Instructing others to do something with an object is an ordinary practice that may involve embodied demonstrations. Based on a single case analysis of a “troublemaker” example we show how the instructor decomposes the general instruction into more mechanic steps that can be mimicked by the instructee. The example is chosen because it reveals some of the taken for granted knowledge involved in accomplishing an instruction. The case is from a video mediated civil service setting in Denmark. In this setting, it is often necessary for a citizen to use a locally present object, e.g. a printer, and this may require instructions from the employee who is physically located elsewhere. We show how the instructor within a sequence decomposes the instruction from indexical references to known practices for dealing with objects to a more simplified discernible step-by-step description orchestrated by mimicable embodied demonstrations. Thus, the paper contributes with new insights about two aspects of recipient design in instructions and embodied interaction specifically linked to a video-mediated setting. The study is based on eight hours of video recordings of video-mediated encounters and applies EMCA multimodal interaction analysis.

Key words: *EMCA, mimicry, multimodality, video mediated interaction, objects, gestures, embodiment, instructions, recipient design,*

The aim of this paper is to establish new insight about recipient design (Sacks et al., 1974; Depperman, 2015) in video-mediated interaction: the practice of decomposing an instruction and using embodied demonstrations to accomplish a local action. By using a single case analysis of a troublesome example from a data corpus collected in Danish citizen service institutions, where citizens video-interact with remote employees, the analysis will reveal two new insights about recipient design in instructions:

- how problems with understanding an instruction is managed by a sequential practice of *decomposing* the instruction into discernable mechanical steps;
- how one important resource in this practice is to upgrade the embodied work and specifically to produce *designedly mimicable embodied demonstrations*.

Thus, this paper contributes with new insights about a type of practice (decomposing) and embedded within this a type of action (mimicable embodied demonstrations) employed by participants to solve emergent problems of understanding in video-mediated interaction.

Instructions are in face-to-face encounters accomplished through multimodal actions, in which speech and the handling of objects is entangled with embodied displays, illustrations, and tactile actions (e.g. De Stefani & Gazin, 2014; Mondada, 2014b; Nishizaka, 2014). By exploring a naturally occurring “troublemaker” example (Garfinkel, 2002, pp. 125-26) from a distinctive interactional video-mediated environment we are able to explore the subtle details of the interactional work and shared knowledge that ordinarily goes into an instruction unnoticedly, and thus to obtain new insight about the practice of decomposing and using embodied demonstrations in instructional sequences.

However, before we analyze the troublesome case in depth we want to just very briefly show a contrast example of how an “uncomplicated” instruction might look like in this video-mediated public environment. A comprehensive description of the data, setting and transcriptions is provided in the next section. Here, it is sufficient to say that a citizen needs help with an official letter. He has brought along a bystander for the meeting who is the instructee in the sequence. This is a crucial difference between this example and the one we will focus on later: the instructee here is not the target group for the public service. Sitting in a public library he is video-interacting with the instructor - an employee from citizen service. The employee instructs the bystander to put the letter in a printer, so that it can be scanned and thereby read by the remote employee. IOR is the instructor to the right. IEE is the instructee to the left. Transcripts follow the Jeffersonian system and a simplified version of Mondada’s conventions for multimodal transcriptions (Mondada, 2014a), see Appendix.

Example 1: That orange tray

Example 1: That orange tray

| | | |
|--------|---|-------------------|
| 1 IOR: | så vil jeg gerne ha at du: lægger den op i# øh den •bakke I would like you: to put it op in uh that tray | |
| IEE: | | •scans l to r---> |
| Fig. | #figl | |

we presume, that shared understanding about language (e.g. syntax, lexical semantics, phonological and prosodic features), objects (what a printer is) and their visual-spatial properties (where a printer could be) are basic required competences. But we want to unfold the argument and show what participants might do, when they do not rely on these shared understandings, and consequently have to decompose the instruction into more mechanical steps and upgrade the use of embodied actions, in particular the use of mimicable demonstrations.

1. The setting and the data

This paper is based on findings from a research project in Denmark concerning the constitutive features of video-mediated interaction. Data is from business settings, medical settings and public service settings. Across different settings, we have collected instances in which part of the encounter consists of recipients being instructed in accomplishing required actions involving object handling, e.g. handling documents, printers, medical equipment, and computers. On the basis of our corpus of video recordings (+100 hours), we have a large corpus of object-related instructional instances. For this article we have chosen only to focus on one specific setting and activity type: Video-mediated encounters from public libraries where citizens seek help with various things regarding the welfare system. Often in the course of these conversations, it becomes relevant to scan an official document that the citizen has brought to the meeting. These citizens have not been able to use other available solutions to their problems because they are, as the official documents state: “not digitally ready”. This is a heterogeneous group as the underlying reasons for being “non-digitally ready” are quite diverse, e.g. illiteracy, lack of language proficiency in Danish or English, various cognitive or physical disabilities, lack of technical know-how, etc. Importantly, the citizens have all been in face to face contact with locally present public service professionals who have deemed them fit to have the video mediated conversations. This means that while participants in our data might have a limited proficiency in Danish, they are by no means aphasic or otherwise communicatively impaired.

Consultations about legal and official texts are in Denmark increasingly replaced by video-mediated consultations, but the activities still involve manipulation of objects and instructions about how to do so. The setting is institutional in the sense that employees from

governmental institutions interact with citizens about institutional agendas. However, we do not focus on the institutionality of the setting as it is not revealed as a member's concern.

Participants are – as shown in example 1 - sitting in front of a computer screen with a webcam integrated in the monitor, in a room designed for video encounters. The instructor is an employee from citizen service. She is sitting at a desk in an office across town, also in front of a computer screen, interacting through a video camera mounted on top of the screen. We have a corpus of 8 examples from this specific setting where citizens are instructed to put a document in the printer.

The specific case we will unfold in this article is from a video-mediated civil service encounter between an employee (instructor (IOR)) and a citizen (instructee (IEE)). IEE is instructed to locate the printer in the room ultimately to be able to scan a document. This rather mundane activity of locating an object is turned into a complex action due to the complicated and ambiguous perceptual fields of the setting, and the employee tries to solve it by decomposing the instruction and upgrading the use of multimodal resources. We show how the situated material context constitutes a frame for the activity and how this is part of IEE's problem with solving the task. Some of these encounters are quite uncomplicated, as shown in example 1, but sometimes the sequence is expanded because of trouble with locating the printer. We believe this case may serve as a troublemaker example, that – as Garfinkel envisioned – may reveal subtle details of the taken-for-granted knowledge in ordinary affairs and the resources involved in accomplishing the task nevertheless.

This single case analysis is chosen because it exhibits a very important insight about instructions: they are composed of taken-for-granted knowledge about language, objects and space which is revealed through repair-initiations and the recipient designed practice of *decomposing* the instruction going from indexical verbal references to a step-by-step instruction using *mimicable embodied demonstrations*.

The case is also chosen because it exhibits some of the “teleepistemological” aspects (Dreyfus, 2000) of video-mediation as a fractured ecology that may complicate the smooth organization of the interaction. The settings are semiotically rich compared to e.g. written instructions, telephone calls and YouTube instructional tutorials, where there are limited possibilities for interaction and inspection, but the settings afford fewer possibilities compared to face-to-face interaction. We will show how the fractured ecology is consequential for the unfolding of the interaction.

2. The fractured ecology of video-mediated interaction

Instructions occur, obviously, in all sorts of settings. We are particularly interested in instructions during video-mediated encounters where participants are virtually co-present and thus restricted by the affordances the video camera provides and produce as a “fractured ecology” (Heath & Luff, 1992; Luff et al., 2003; Luff et al., 2013; Luff et al., 2016). Instructions are particularly interesting to study in a video-mediated environment because the appearance of a visual action in the local environment is different when translocated to the remote environment. EMCA studies of embodiment and multimodality reveal how referential activities, deixis, ‘pointing’, and so on are situated, collaboratively achieved, and shaped moment by moment by participants in the course of their face-to-face interactions (e.g. Nevile, 2015). But to point or use deictic terms in a fractured ecology, where the screen inverts images and dislocates the reference, is potentially complicated.

Video-mediated interactions have been explored with respect to activities as diverse as remote supervision of surgery (Mondada, 2007b), teaching sign language (Hjulstad 2016), international business meetings (Author forthc. A), the use of telepresence robots (Due, in press) and project group work mediated by a telepresence system (Luff et al., 2016). A number of studies have focused on the interactional implications of technical properties, i.e. that screens are usually small and project only a selection of what is visually available at the connected locations (Raudaskoski, 1999), that interaction may thus be affected by reduced visibility as well as delays caused by technical difficulties, that there is a lack of haptic experience (Denstadli, Julsrud, & Hjorthol, 2011), and that participants may compensate for reduced visibility by e.g. verbalizing actions (Author, in press). Of specific relevance for our study is Licoppe and Veyrier (2017) who have studied the ways in which participants are shown on the screen in multilingual courtrooms and in which participants show objects to each other in video-mediated collaborative settings. Licoppe (2017) and Licoppe et al. (2017) zoom in on the practices of showing objects; how they are held and manipulated. We contribute to these studies by showing how objects are not just showed and how gestures are not just produced as e.g. pointing practices, but how instructors also may design embodied demonstrations to be mimicked by the instructee in order to for him/her to see or handle objects correctly.

3. The practice of decomposing and retreating to embodied demonstrations

Instructing somebody is an accountable and a paired action (Garfinkel 1967; 2002 chp 6; Mondada 2011) that creates local identities of instructee and instructor, and induces a questioning of competence to the interaction, as well as entitlement to a normative stance (Garfinkel 1967; Macbeth 2011). Instructing others retrospectively constructs something prior as repairable. Instructions may be firmly rooted in institutional roles and their distributed rights and obligations. Exploring instructions thus raises the questions of how instructions are collaboratively built, designed, situated and made sense of in a course of action. In this paper we aim to understand how instructions to use objects are accomplished in troublesome case. Studies of face to face instructions and demonstrations specifically related to the use of objects includes Goodwin's studies on the use of a Munsell chart (Goodwin, 1993) and how instructors during anthropological fieldwork achieve a public event that can guide the perception of the recipient in the course of seeing (Goodwin, 1994). Goodwin introduces the notion of "highlighting" as a member's practice for steering the attention of the co-participant, partly also by the use of objects. But the instruction may be insufficient and the instructor may have to decompose the instruction into discernable steps.

We take the term *decomposing* from the work of Goodwin (2013) who describes how many different kinds of resources can be selectively decomposed, reused and transformed to build a next action. In later studies of human embodied actions Streeck (2017) has shown how the work of 'explicating' objects is to indexically or figuratively decompose them, that is, to exhibit them as things made of separate parts. Jan Svennevig has shown how people usually try the most economic and therefore easiest solution first (Svennevig, 2008); e.g. to refer minimally to things (Sacks & Schegloff 1979), instruct minimally and hope that the indexical references and shared knowledge will do the work. Recently, Jan Svennevig (2018) has shown in an experimental study, where collaborating informants had no visual access to each other and no shared physical ecologies, how the instructee's understanding of instructions increase if the verbal instructions are decomposed into small main clause increments. Thus, decomposing is participants' work of disassembling a sequence into smaller units.

We will argue that one specific type of action within the sequences of decomposing is the instructor's production of mimicable embodied demonstrations. Recent

research has shed light on embodied demonstrations from face to face settings. Keevallik (2013) has e.g. shown how bodily demonstrations in dance classes are syntactically linked to the usually incomplete structure of a sentence, which projects a bodily demonstration to be a next action. She shows the transitional work being done from syntax to continuation realized by the body. Hindmarsh, Hyland and Banerjee (2014) have shown how instructors in pre-clinical dental education shifts from verbal instructions to also demonstrating the required practice by using the relevant objects. Evans and Reynolds (2016) have shown how sports coaches propose a correction in power lifting and basketball sessions with an embodied demonstration. Due have shown how people in creative business meetings use embodied demonstrations and illustrations as a resource for co-constructing an imagination space (Due, 2016). And Nishizaka (2017) has shown the work applied in structuring bodies in such a way that they can be viewed as analogous during a demonstrative act in prenatal examinations. We contribute to this line of research in embodied demonstrations by providing a systematic analysis of a) how the action is embedded in a decomposing instructional sequence, b) how the video-mediated fractured ecology constitutes a different framework than face to face, and c) how the demonstration is designed to be mimicked.

4. Analysis: The structural evolution of instructional actions

We will be looking at the opening of the instructional sequence as it is, for the participants, a matter of ‘getting to first base’. Leading up to the excerpt, IOR and IEE have established that IEE needs help to understand the content of a letter she has received from the authorities and which she has brought with her. This activity is verbalized by participants as the *reason for the encounter* (cf. Drew & Heritage, 1992). IEE is not sure who sent her the letter. In order to help IEE, IOR wants IEE to scan the letter in a printer physically located to the right of IEE and the computer screen – just as in example 1. To do this, IOR needs IEE to shift her visual focus from the screen to the printer, thereby creating a new participation framework for the next step of the activity. The analysis of the single cases is cut into analytical sections in order to highlight the stepwise progression, moving from the minimally designed verbal instruction to the expanded decomposing sequence where the instruction is bodily demonstrated.

4.1 Failures of indexical reference to “common knowledge”

We enter the situation as IOR tries to mobilize a shift in the activity by instructing IEE to turn her head by indicating the location of the scanner in the room through indexical references.

Example 2: An orange tray

1 IOR: Δka du Δse ovre til højre for digΔ der står der en
 can you see over to your right there is a
 Δ.....Δpoints----->Δ
 2 printer#,
 printer
 fig #fig.1

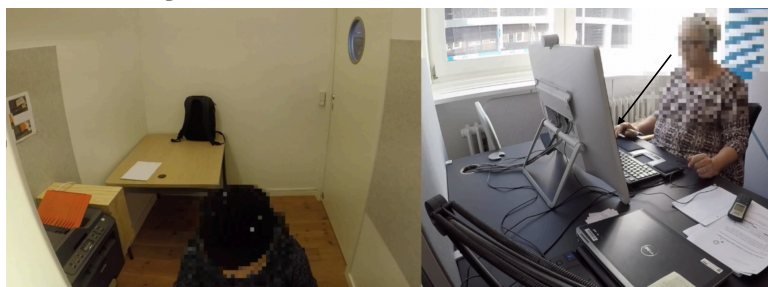


figure 1

3 (.)
 4 IOR: med en orange #bakke på.
 with an orange tray on it
 fig #fig.2

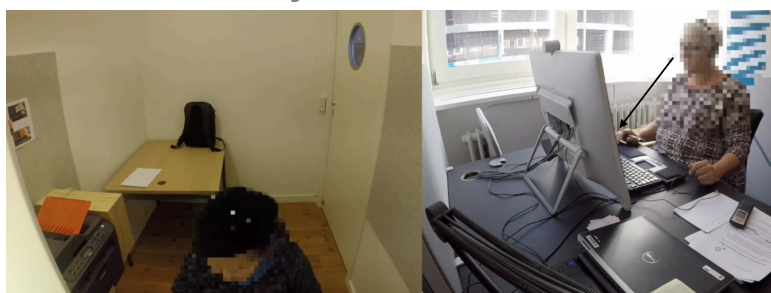


figure 2

5 (1.0)
 6 IOR: til højre for dig•
 to your right
 looks down --->•
 7 IEE: •*>den der<•
 >that one<
 taps paper
 •up from paper
 8 (0.7)
 9 IEE: >dender<=
 >thatone<
 10 IOR: =Δnej til højreΔ [ik] lige foran
 no to the right [not] straight ahead
 Δpoints at scrΔ
 11 IEE: [*mhm*]
 [*mhm*]
 -l- (1: raises r hand)
 12 (0.8)

In line 1 - 2 IOR constructs a first attempt to instruct IEE to locate the scanner on her right. The instruction is formatted as a yes-no interrogative inquiring about IEE's ability to see the artefact. There are two things we would like to draw attention to: First, for reasons that will become apparent later in the analysis, the lexical choice of "printer" (l. 2) to refer to the artefact might be problematic in terms of recipient design. Secondly, we observe that IOR simultaneously to her linguistic action points to the visual field where the scanner is represented on her own screen, a gesture that cannot be indexically meaningful in the interaction due to the fractured ecologies of the setting (cf. Luff et al., 2003). During the initial instruction, IEE does not demonstrably show any uptake to a restructuring of her body posture, as she sits still with her gaze directed towards the letter she is holding (fig 1). In figure 1, we can also observe that IOR's gaze is simultaneously directed at her screen, i.e. the visual field where she can monitor the lack of bodily uptake by IEE.

After a beat of silence, IOR adds a syntactic increment to her description ('with an orange tray on it', l. 4), orienting to the lack of uptake to her instruction. The increment deals very directly with the object's salience in a complicated perceptual field: the increment is *highlighting* (Goodwin 1994) in the sense that it is designed to steer the attention of the co-participant towards an object in the perceptual field, i.e. "*an orange tray*", l. 4. As an ethnographic addition to this, the orange tray can be seen as a "historically sedimented resource" (Goodwin 2018) for the participants, as the orange colour has been added by the public service institution to make it more salient. In other words, part of the material surround has been highlighted in a very literal sense as a remedy to the recurrent interactional trouble of the professionals to successfully instruct the citizens to locate the scanner. During the increment, IEE shifts her gaze from side to side, still facing down towards the letter she is holding (fig 2), a posture not optimal for her to be able to see the scanner. IEE keeps her posture and focus during the 1.0 second pause in line 5, thus still failing to comply with the instruction.

IOR now goes on to reformulate her initial instruction and focuses on the lack of reorientation by IEE ("*to your right*", line 6), omitting other information and narrowing in to a directional guidance. Following this reproduction and focus on orientation, IEE looks up at the screen, and says "*that one*" (line 7), while she taps the letter twice, rhythmically tying the gestures to "*that*" and "*one*", thereby producing an embedded interrogative declaratively

intoned. By doing this she shows lack of understanding of IOR's turn, as she does not reorient her focus from the initial participation framework of the activity (which would be the result of following the instruction).

IOR does not initially follow up on IEE's turn, and following a pause IEE reproduces her turn in line 9 with the same wording delivered in a faster manner ("*>thatone<*", l. 9), looking at the screen with her eyebrows raised. In line 10 IOR responds to the first pair part with a latched turn, beginning with "no" followed up with a specification of the direction of C's intended focus in the room reusing a minimal reformulation of the direction "*to the right*" and adding "*[not] straight ahead*", citing IEE's current body posture and orientation as incorrect.

Thus, we observe in the data several issues of problems with indexical references to common knowledge. The mere verbal reference to a printer in the room is not enabling IEE to do the required action. Now, this prompts IOR to decompose the instruction, as we will see in the next section.

4.2 Decomposing gloss to verbalized step-by-step instruction

As the descriptions of the printer's position did not succeed, a slot for more detailed instruction is produced, and the instructor resorts to step-by-step verbal action-directed instructions:

```

12      (0.8)
13 IOR:  Δ.hh Δprøv å drej  ΔhovedetΔ
          .hh try to turn the head
          Δ....Δswings pen---Δ,,,,Δ
14      (.)
15 IOR:  til højre
          to your right
16      *(1.0)*(0.8)*
          *.....*--2--*
17 IOR:  drej dit ΔhovedΔ
          turn your head
          Δ--3--Δ
18      (0.6)
19 IEE:  mit ho::
          my head

```

2: turns paper to face screen
3: twirls index finger

After the failed attempt to instruct IEE in the preceding lines, IOR changes the linguistic packaging of the instruction in line 13 and produces an imperatively formatted instruction (*.hh try to turn the head*). By this formulation, she verbalises *one* crucial component in the required course of action that IEE is to undertake if she is to locate the scanner, thereby decomposing the instructional gloss (“*can you see over to your right there is a printer*” l. 1 – 2) into an explicit head movement: *turn the head* (l. 13). IOR recycles her instruction a couple of times (l. 15 - 17), as IEE does not comply with the instructions but instead shows an active orientation toward the letter she is holding by turning it up toward the screen as if to show it to IOR (l. 16). After a pause in line 18, IEE repeats a phrase from the prior utterance, perhaps to solicit an other-initiated repair or elaboration, and in any event demonstrating a lack of understanding of the instruction. She still has not turned her head and accountably seen the scanner to her right. In the excerpt, we see that IOR upgrades her use of multimodal resources: she uses different hand gestures (l. 13 and l. 17), even though these are not necessarily in IEE’s field of vision. We also see an upgrading verbal mode from interrogative with embedded declarative (l. 1) to increments to declaratives (l. 4 and 6), other-correction (l. 10) and mitigated directive (l. 13), finishing with an unmitigated and minimal imperative: “*turn your head*” (l. 17). The upgraded use of multimodal resources however does still not lead to a successful outcome, as we will see in the next excerpt.

4.3 *Decomposing* instruction into discernible mimicable steps

As IOR's instructions have proven unsuccessful this far, she now decomposes her instruction into a discernible mimicable action, using part of her own *body as a model* (Nishizaka, 2017) for IEE:

```

20 IOR:      Δsā- #sādan her*                Δ*
             li- like this
            Δturns head back and forth---Δ
                    *-----4-----*
            ((C also turns head, still looking at screen))
fig          #fig.3

```

4: turns head to the right, raises right hand

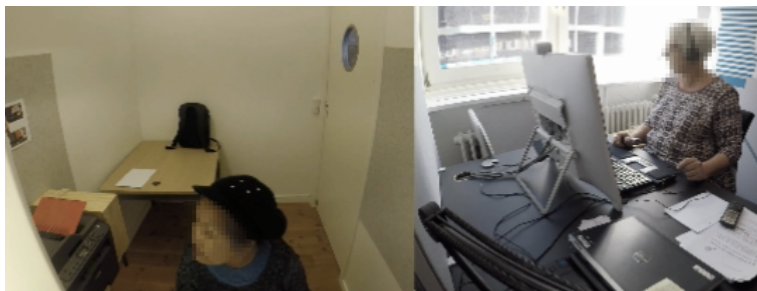


figure 3

- 21 IEE: det høj-
the righ-
- 22 IOR: Δden vej jaΔ (.) kig så
that way yes (.) now look
Δ.....Δpoints--->
- 23 •(0.7) •
•looks down•
- 24 -38 [...] (IOR keeps saying to IEE that she must turn her head)
- 39 IOR: Δgør så*dan her
do like this
Δturns head---> (until l. 52)
*turns head---> (until end of transcript)
•to the right ---> (until l. 42)
- 40 IOR: ja
Yes
- 41 (.)

In line 20, IOR now bodily demonstrates the head movement for IEE: she shakes her head from side to side while producing “*li- like this*”. This shift to a bodily demonstration by IOR is occasioned by the interactional contingency of the situation, that is on the failure of IEE to follow the instructions in the preceding interaction. The demonstration is designedly mimicable as the deictic terms (“*li- like this*”) are only intelligible in the entire temporal and spatial body configuration, referring to IOR’s head movements. It is worth noting that while turning her head, IOR keeps her own gaze fixed on the screen, possibly monitoring IEE’s actions which is an integral part of the instruction activity – and actually mirroring and

implicitly legitimizing the same (problematic, yet obviously relevant) actions by IEE. By maintaining her gaze at the place where she can observe her co-participant, IOR thereby embeds the demonstration in the overall instructional activity (See Nishizaka 2017: 114 for a similar argument).

At the beginning of IOR's second head shake in line 20, IEE turns her head and holds it to the right, while keeping her gaze focused on the screen (Figure 3). While this head position complies with the formal instruction (turning her head), it does not do so in a manner that leads to the intended outcome, i.e. that IEE actually switches her visual field in a way so she sees the scanner in a relevant, accountable way. IOR can observe this by the fact that IEE's gaze has not moved to the side but remains fixed on the screen. In the next turns, IEE displays her understanding of the term "right" by holding her right hand up and saying "the right" (l. 21) but does not display the scanner as a relevant thing on which to focus. IEE demonstratively orients her gaze towards the screen and thus fails to locate the scanner. IOR continues instructing and showing IEE to turn her head (omitted in the transcript), until she recycles her *designedly mimicable embodied demonstration* of the required action in l. 39 (turning her head and saying "do like this" simultaneously). Again, IEE complies with the instruction as she turns her head, but keeps gazing at IOR and thus does not accountably recognize the scanner as relevant. She acknowledges the instruction with a "yes" in Line 40 but does not produce a change of state token or similar item. Integrated with these embodied moves, the verbal mode is upgraded from minimal deictic references (l. 20 and 22) to minimal unmitigated imperatives (l. 22 and 39).

Thus, this sequence shows how IOR decomposes her instruction into an embodied demonstration of intermediate steps which is designed to be mimicked by IEE, and that IEE subsequently mimics the actions. However, the project of getting IEE to see the scanner as a relevant object still does not succeed. IEE does in fact exactly mimic IOR's actions; IOR turns her head but do not change gaze direction. IEE mimics this by turning her head slightly but keeps gazing at the screen. So, although the actions are designed to be mimicked, and are in fact mimicked, they still do not solve the local, emergent problem. The 'fractured ecologies' of the interaction seem consequential for this, as the technological setup only allows the participants to mutually monitor each other's embodied behavior if they focus their gaze on the screen that renders visible parts of their co-participant's ecology. In a co-present demonstration of mutual actions this mutual monitoring would not be restricted

to a software window on a screen (see e.g. Goodwin 1994, Keevallik 2013, Nishizaka 2017).

4.4 Upgrading the use of multimodal resources to make actions mimicable

Given that the instruction remains unsuccessful and that IEE does not demonstrate understanding of what to do, we observe how IOR upgrades her efforts to produce the instruction through deictic pointing practices and by demonstrating and designing mimicable actions.

```
42 IOR:      •Δ↑>DER<#  
              ↑>THERE<  
              Δpoints--->  
              --->• at screen --->
```

fig #fig.4

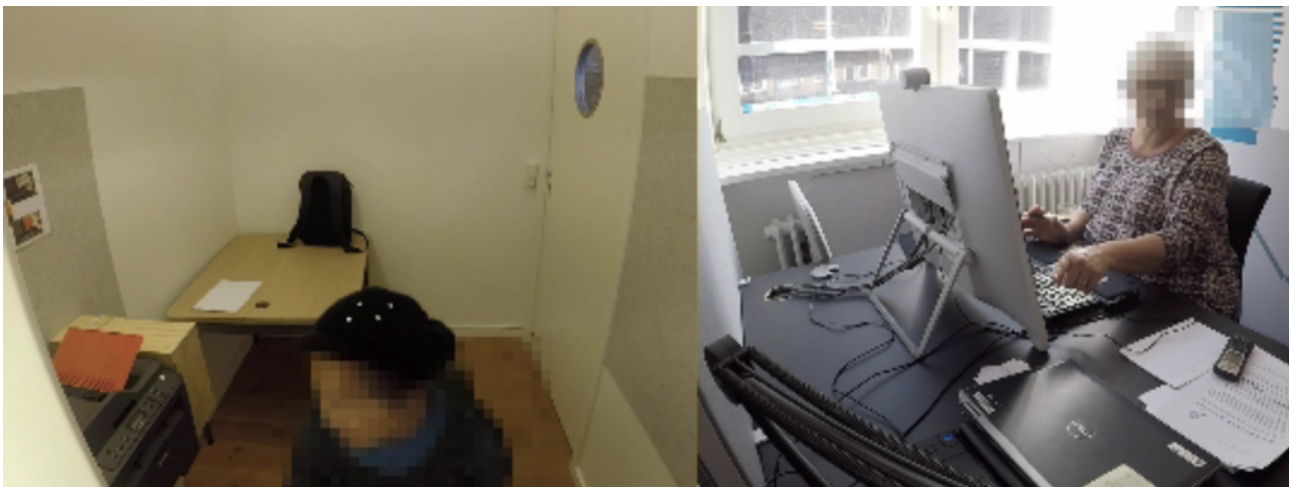


figure 4

```
43          (0.5)  
44 IOR:      der står en maski:n,  
              there is a machi:ne,  
45          (0.4)  
46 IOR:      ja  
              yes  
47          (.)
```

48 IOR: med en orange bakke
with an orange tray
 49 (0.4)
 50 IOR: ja
yes
 51 (.)
 52 IOR: prøvΔ- drej hovedet Δlidt mere
try- turn your head a bit more
 -->Δ Δturns head more to the side --->
 53 (0.3)
 54 IEE: ja
yes
 -->•
 55 (0.9) ((C sees printer))
 56 IOR: [↑DER#Δ
[↑THERE
 --->Δ
 •looks at printer•
 fig #fig.5
 57 IEE: *[den der *
[that one
 ---5---

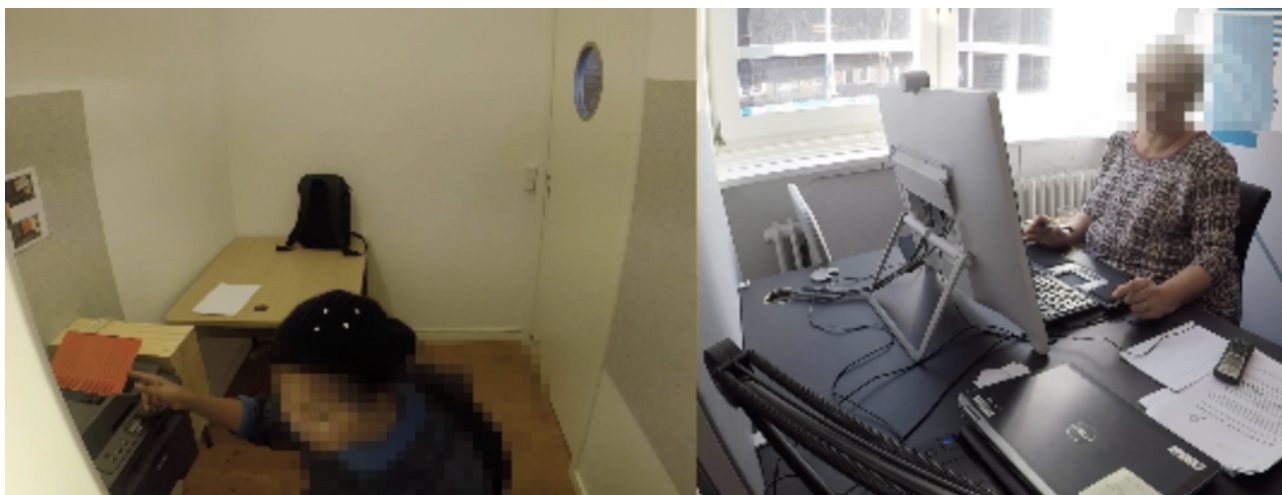


figure 5

58 (0.5)
 59 IEE: >den der<=
>that one<=

```

60 IOR:    =ja
           =yes
           -->•
61         (0.7)
           -->Δ
62 IOR:    put papiret derop
           put the paper in there

5: points at printer

```

As both IOR and IEE are holding their heads to the side, IOR produces the demonstrative pronoun with high volume “↑>THERE<”, creating a deictic reference for IEE to make sense of. The timing and the format of the reference are crucial as it is produced at a time when IEE is holding her head to the right and is observably gazing to the right as well. The rapid production, minimal design, and high pitch of the reference stresses its deictic nature: It is upon the production of the reference that something in the visual field of IEE is indexed as urgently relevant. Interestingly, IOR also points to the scanner on her own screen on l. 42, but this is not effective for the interaction, firstly because the camera cannot pick up this gesture at the height at which IOR is holding her hand and secondly because the pointing is meaningless because *this* recipient has severe difficulty with understanding and recognizing the terms and object involved in the instructions, partly due to the fractured ecology of the encounter. As with the previous instructions, this instruction only partially succeeds, since IEE does not see the printer. Following a short pause, while IOR and IEE are still holding their heads slightly sideways, IOR follows up her embodied instruction with “there is a machi:ne” (l. 44). This turn is a modified repeat of her instruction in l. 1-2. The modified element is the reference term for the object IOR is trying to get IEE to see: IOR switches from “printer” to “machine”. This switch presupposes that IEE has less knowledge of the object than the initial reference, as “machine” is the hypernym of “printer”, and in so doing targets the initial formulation of the instruction as potentially problematic.

IOR’s attempts to get IEE to locate the scanner are unsuccessful until line 52. Here, she elaborates her bodily demonstration by turning her head (which is already held in a sideways position) a little further to the side as she says, “a little more”. IEE acknowledges this instruction verbally and by turning her head even more to the side. As she does so, she makes a subtle movement with her head, which somehow indicates a change in perception,

i.e. seeing the printer. As IEE turns her head all the way to the side and moves it subtly upon seeing the printer, IOR repeats her deictic reference from earlier (↑ *THERE*, l. 42). In line 57, just as IOR has finished her turn, IEE points to the scanner and adds to this gesture verbally with the minimal question designed declarative: “that one” (l. 57).

Thus, the instruction and reorganisation of the co-participant's body posture and perceptual field is accomplished through a mix of semiotic means: both a verbal instruction and designedly mimicable embodied demonstrations of moving or turning one's head, using the body as a model, even more following a respecification of the referent to “machine”, which may also play a part in the success of the instruction.

The problem is not that IEE does not mimic the embodied actions; the problem is that the gaze is not moved away from the screen. IOR turns her head (e.g. l. 39) but keeps gazing at the screen, and IEE mimics exactly that action. This reveals interesting aspects of visibility. Not only in terms of what to see and how to see it and in terms of resources involved (moving head and/or shifting gaze), but also with regards to comprehension expressed by use of shape and color terms to direct IOR's actions. As part of the upgrading sequence IEE breaks down the instruction into smaller parts consisting of reference to ‘sense-data’; a vernacular use of basic shape, positional, and color terms to direct IOR's search.

It is interesting how the construction of the perceptual field is reconfigured from indexical references to presumed common knowledge about what a printer is, to its decomposed elements of being a machine with an orange tray. The construction of seeing something as something is a very basic semiotic task involving an interpreter seeing an object as a sign (Peirce 1955). Already Hanson (1977, p. 15) showed how we do not see things the same way. Looking at an X-ray tube, Sir Lawrence Bragg and an Eskimo baby would visually see the same object. But the ways in which they are visually aware are profoundly different. Seeing is not only the having of a visual experience; it is also the way in which the visual experience is had. In the example IEE tells IOR to see “the printer” as a machine with an orange tray, thereby decomposing the instruction into more basic elements. Seeing is an organizational feature of an embodied, visible activity. Visual perception is not simply a discrete state resulting from information processing, but rather a resource for the activity of information processing (Nishizaka 2000). We do not know if the recipient of the instructions is unfamiliar with what a printer is, but we observe the practice of decomposing

and in that process IOR reformulates the printer as a machine. So, IOR orients to the possibility, that IEE cannot really “see” the printer, as she may not know what a printer is. “Seeing” is thus not only a matter of establishing a common visual field by gazing the right direction, which may be more complicated in the video-mediated environment, but also a matter of seeing something as something, i.e. understanding (Cf. Coulter and Parsons, 1990).

5. In the absence of shared understandings: Mimicable embodied demonstrations in decomposed sequences.

The fundamental order of human action relies on tacit norms (Wittgenstein, 1953; Garfinkel, 2002; Liberman, 2013) which is very useful for the seamless accomplishment of activities. Research has shown that a successful interpretation, an appropriate expression, and the practical application of general social rules do not only follow the simplicity of words, but involve degrees of situated appropriateness, implicit knowledge which is neither sufficiently decontextualizable, nor capable of being converted into explicit knowledge (Dreyfus & Dreyfus, 1986; Merleau-Ponty, 2002; Loenhoff, 2017). In this paper we have shown that the accomplishment of a rather simple task of being instructed to put a paper in a printer is composed of taken-for-granted shared understandings embedded in language-understanding (unable to naturally read a letter), object-understanding (knowing what a printer is) and a visual-spatial understanding (knowing where a printer could be placed) that is situated within a particular relationship to the local, emergent circumstances (cf. Suchman, 1987). This is made evident when the interaction is troubled and sequentially decomposed.

The analysis has revealed how complicated actions become when instructees do not understand and act according to implicit understandings and practical knowledge. The instructee did demonstratively not have an intuitive grasp of the situation based on tacit understanding of the activity. But this lack of expertise in exactly this kind of situation was solved through an adjusted recipient design, decomposing the required action into simple discernible steps, specifically by upgrading the verbal turn design and using the designedly mimicable embodied demonstrations. By “solved” we mean that the instructee finally was able to recognize the machine (printer) as the place to put her paper. In the subsequent sequences, which we have not included in this analysis, she is actually able to put the paper

in the printer without complications. The instructor then controls the scanning process from her remote position, thus reducing any further complications on the recipient side.

We find similarities to James Heap's (2014) discussions on reading lessons in elementary school, where assessments of oral reading competency use a distinction between reading with and reading without “understanding”. We observe the same kind of issues in this completely different setting. An instructee can perform the required handling of an object by “fully understanding” what to do, how to do it and why to do it. This requires “only” that the instructor says for instance: “put the paper in the printer” and then – based on prior knowledge and expertise, the instructee scans the room and recognizes the printer (as shown in ex 1). But the required action can also be accomplished – as we have shown in the longer single case - while not really “understanding” how and why, but just by mimicking discernible steps.

Studies have shown how children from early ages learn by imitating other people (e.g. Byrne & Russon, 1998; Bekkering, Wohlschläger, & Gattis, 2000; Want & Harris, 2002). However, to mimic and to imitate are not exactly the same. While imitation may mean to do something in the same way as something or someone else, mimicry is a more mechanical type of behavioral copy of someone's actions or other observable expressions. Thus, to mimic may have more of a connotation of “super-copying” (McGuigan, Makinson, & Whiten, 2011) of another individual, as though one is copying the outward actions of another but not necessarily understanding why those actions are done. This is perhaps why mimicry also have been applied in robotics, where robots are programmed to perform certain actions. In this context, it is necessary for the robot (being purely algorithmic) to decode clear-cut steps of an action or activity in mechanical ways (Alač, 2009). We have shown just how in the face of interactional indications of a lack of ‘understanding’, the instructor decomposes the interaction and resorts to a more mechanical or behavioristic sequence of prompts and bodily demonstrations designed to be mimicked.

In our data, we observed an upgrade of the verbal mode and a move towards minimization and granulation of explanation. The instructor moves from using interrogatives with embedded declaratives or inquiries about abilities, to declaratives, deictic references and increments to declaratives, and finally resorts to unmitigated imperatives, as recorded in the following list of upgrading attempts:

- from references to common knowledge about objects to less precise references to unfamiliar objects (e.g. l. 2 versus l. 44)
- from references to common knowledge to use of under-specified discourse referents, e.g. the demonstrative "this here", a descriptive use of indexicals (cf. Nunberg, 2004), (e.g. l. 1-4 versus l. 20 "like this" and l. 22 "that way" and l. 39 "like this")
- from requests with references to common knowledge about objects to directives with a granularization of context (cf. Nunberg, 2004), (e.g. l. 1-4 versus l. 6 versus l. 10)
- from inquiries and mitigated verbal directions to unmitigated imperatives (e.g. l. 1-2 versus l. 13 versus l. 17)
- from requests to directives (e.g. l. 1 versus l. 13)
- from using the verbal mode to using embodied demonstrations (e.g. l. 1-4 versus l. 13-22 versus l. 39-52)
- from showing physical actions to stylized or emphasized demonstration of physical action (e.g. l. 1 versus l. 13 versus l. 20 versus l. 39)
- from embodied gloss of movements to embodied decomposed granulated physical actions (e.g. l. 1 versus l. 39, 42, 44, 48 versus l. 52, 56, 62)

The upgraded versions all treat the lack of desired uptake that IOR is pursuing from IEE by means of physical actions with respect to the scanner as indications of the IEE not understanding or not seeing, not as the IEE not being willing or able to do so. The apparent complexity and difficulty of this task is about lack of shared language-understanding, object-understanding and a visual-spatial understanding that is situated within a particular relationship to the local, emergent circumstances but it is also related to problems based on the fractured ecology of the video mediation.

One specific characteristic of video-mediated interaction that contributes to the complexity of the instruction activity is the fact that the participants do not share the situated material structure in which the activity is being conducted. Therefore, the instructors cannot use tactile resources or other locally situated sensory input such as smelling/feeling/sensing when assisting the activity and consequently has to instruct the instructee on how to handle the objects themselves in a remote place by using verbal descriptions and embodied demonstrations. This fact distinguishes the setting we have analyzed significantly from the

studies of face-to-face demonstrations by e.g. Goodwin (1994), Keevalik (2013) and Nishizaka (2017).

Although the single case analysis showed how basic communication is so fraught that the relevance of the video is swamped by difficulties with understanding, the video-mediation is nevertheless a fundamental precondition that restricts the possible repertoire of actions with procedural consequentialities. The video-mediated format and the specific restrictions and affordances it provides produce structural problems with regard to locally meaningful instructions (e.g. where a pointing action is aiming at). This is not a new finding, but it adds to the wider literature concerning tele-medicine (Jarvis-Selinger, Chan, Payne, Plohman, & Ho, 2008; Liu et al., 2007; Miller, 2003, 2011; Stryhn, Jackson, & Nielsen, 2016) as well as, more specifically, to the EMCA research done in this field (Heath & Luff, 1992; Luff et al., 2003; Luff et al., 2013, 2016; Arminen, Licoppe, & Spagnolli, 2016; Hjulstad, 2016; Licoppe et al., 2017; Licoppe & Veyrier, 2017).

Licoppe et al. (2017) have demonstrated how participants show one another objects in video-mediated settings, revealing how this conduct is configured and organized within the interaction between participants. They and Author (in press) show how the subtle adjustment of bodily position, head, and gaze with respect to the handheld objects offer crucial resources for participants seeking to achieve joint seeing and shared understanding. We have in this paper likewise shown how participants readjust their bodies in space in order to render actions visible (cf. Author X). However, unlike Licoppe et al., we have shown just how instructional object-directed actions are upgraded when shared understanding is problematized. Thus, this paper contributes with two new findings about the *practice of decomposing* and the action of doing a *mimicable embodied demonstration* as participants resources for succeeding with the current task in a video-mediated environment.

We showed how the embodied demonstrations are designed to be mimicked, but they require that the instructee performs a visual analysis of the embodied demonstrations. In the example, the instructor turns her head to the right, thereby demonstrating the movement she wants the instructee to make and she also points. These demonstrations are produced by the instructor to actively reduce contingencies so that the instructee can mimic intermediate steps in order to accomplish the activity. The bodily actions of the instructor are not meant to solve things on their 'own side' but solely serve to demonstrate for the participant on 'the other side'.

There is a vast body of research on how participants use pointing gestures to orchestrate verbal actions, specifically with regard to referential, indexical practices and deictics (Goodwin, 2003; Haviland, 2000; Kendon, 2005; Mondada, 2007; Streeck, 2010). However, in the excerpt the instructor's pointing practice is toward her own screen, aimed at something positioned at the instructee's remote ecology: the pointing toward the scanner (l. 42) makes sense for the instructor based on her visual orientation but is not treated the same way by the instructee. This poses a challenge caused by the divergent relationship between what the professional is pointing toward "through the screen" and the way the point appears on the instructees side of the screen. Because of this discrepancy, the pointing activity does not support the communicative purpose of attaining a shared focus between participants.

Additionally, spatial terms (Nishizaka, 2017) were used by the participants to describe position and directions (e.g. line 1: "to your right there"). These kinds of terms are indexical and dependent on reference points in space. But as showed in the analysis, this is complicated in a fractured ecology if participants do not rely on other taken-for-granted understandings. Thus, the instructor deals with the lack of understanding by eventually demonstrating, as part of the instructional sequence, the required actions on herself, using her "body as a model" (Nishizaka, 2017). This is a kind of "phenomenological displacement" (Henry, 2008), where the instructor imaginatively places herself in the position of the instructee and directs the head and arm movements to open up a visual "phenomenal field" (Garfinkel & Livingston, 2003), within which the printer can then be shown to be present. The instructor produces embodied demonstrations by using her own body to make the instructee perform the required physical movement to handle the object correctly. This is different than producing depicting gestures (Streeck, 2009), which are designed to almost literally reassemble an object. Using the body as a model for demonstration in order to be mimicked is a different action but similar conceptually connected to vision – but vision as it is enabled through the local affordances of the video cameras, angles and screens which obscures the image proximity and proportions.

The single case analysis has revealed how troubles in a video-mediated institutional context where the instructee exhibit basic problems with understanding, are solved through mimicable embodied demonstrations in a decomposed sequence. The basic troubles with understanding reveal the indexical knowledge that ordinarily goes into shared

understanding. This makes us think of Wittgenstein’s discussions of how to understand a sign-post. He reflects: “... where is it said which way I am to follow it; whether in the direction of its finger or (e.g.) in the opposite one?” [Wittgenstein 1953 §85]. We have shown how in the context of limited shared taken-for-granted knowledge about language, objects and visual-spatial properties, the instructor competently decomposes the sequence into mechanical granulated steps in the process of accomplishing the activity and make it salient for the recipient just how to “understand” the “sign-post”. This is truly astonishing interactional work.

| APPENDIX | |
|---|--|
| Transcription conventions | |
| Symbol definition | |
| Participants: Professional (P), Citizen (C) | |
| (2.5) | Approximate length of a pause in seconds |
| [But] | Overlapping utterances |
| ↑ | Rising intonation |
| : | Sound stretching |
| = | Latched utterances |
| <u>Stress</u> | Stressed word |
| °ssh° | Low-volume talk |
| * | Gestures done by C |
| • | Gaze done by C |
| Δ | Getures done by P |
| † | Gaze done by P |
| --- | Action’s apex is reached and maintained |
| | Action’s preparation |
| //// | Embodied action is withdrawing, retracting. |
| --->> | The action described continues after the excerpt’s end |
| Fig | The exact moment at which a screen shot has been taken |
| # | A figure’s position within turn at talk |

References

- Author (forthc). X
- Alač, M. (2009). Moving Android: On Social Robots and Body-in-Interaction. *Social Studies of Science*. <https://doi.org/10.1177/0306312709103476>
- Arminen, I., Licoppe, C., & Spagnolli, A. (2016). Respecifying Mediated Interaction. *Research on Language and Social Interaction*, 49(4), 290–309. <https://doi.org/10.1080/08351813.2016.1234614>
- Bekkering, H., Wohlschlager, A., & Gattis, M. (2000). Imitation of Gestures in Children is Goal-directed. *The Quarterly Journal of Experimental Psychology Section A*, 53(1), 153–164. <https://doi.org/10.1080/713755872>

- Byrne, R. W., & Russon, A. E. (1998). Learning by imitation: A hierarchical approach. *Behavioral and Brain Sciences*, 21(5), 667–684.
- Coulter, J., & Parsons, E. D. (1990). The praxiology of perception: Visual orientations and practical action. *Inquiry*, 33(3), 251–272. <https://doi.org/10.1080/00201749008602223>
- Depperman, A. (2015). When recipient design fails: Egocentric turn-design of instructions in driving school lessons leading to breakdowns of intersubjectivity. *Gesprächsforschung - Online-Zeitschrift zur verbalen Interaktion*, nr 16.
- Denstadli, J. M., Julsrud, T. E., & Hjorthol, R. J. (2011). Videoconferencing as a Mode of Communication: A Comparative Study of the Use of Videoconferencing and Face-to-Face Meetings. *Journal of Business and Technical Communication*, 26(1), 65–91.
- De Stefani, E., & Gazin, A.-D. (2014). Instructional sequences in driving lessons: Mobile participants and the temporal and sequential organization of actions. *Journal of Pragmatics*, 65, 63–79. <https://doi.org/10.1016/j.pragma.2013.08.020>
- Drew, P., & Heritage, J. (1992). *Talk at Work: Interaction in Institutional Settings*. Cambridge: Cambridge University Press.
- Dreyfus, H. (2000). Telepistemology: Descartes' Last Stand. In Ken Goldberg (ed.) *The robot in the garden: telerobotics and telepistemology in the age of the Internet*. Cambridge Mass.: MIT Press.
- Dreyfus, H., & Dreyfus, S. (1986). *Mind over machine: the power of human intuition and expertise in the era of the computer*. New York: Free Press.
- Due, B. L. (2017). *Multimodal interaktionsanalyse og videoetnografisk dataindsamling*. Samfundslitteratur.
- Due, B. L. (forth.). RoboDoc: semiotic resources for achieving face-to-screenface formation with a telepresence robot. *Semiotica*.
- Due, B. L. (2016). Co-constructed imagination space: A multimodal analysis of the interactional accomplishment of imagination during idea-development meetings. *CoDesign*, 14(3), 1–17. <https://doi.org/10.1080/15710882.2016.1263668>
- Evans, B., & Reynolds, E. (2016). The Organization of Corrective Demonstrations Using Embodied Action in Sports Coaching Feedback: Corrective Demonstrations in Sports Coaching. *Symbolic Interaction*, 39(4), 525–556. <https://doi.org/10.1002/symb.255>
- Garfinkel, H. (1967). *Studies in Ethnomethodology*. Englewood Cliffs, N. J.
- Garfinkel, H. (2002). *Ethnomethodology's program: working out Durkheim's aphorism*. Lanham Md.: Rowman & Littlefield Publishers.
- Garfinkel, H., & Livingston, E. (2003). Phenomenal field properties of order in formatted queues and their neglected standing in the current situation of inquiry. *Visual Studies*, 18(1), 21–28. <https://doi.org/10.1080/147258603200010029>
- Goodwin, C. (1993). The Blackness of Black. Color Categories as Situated Practice. In L. B. Resnick, R. Säljö, C. Pontecorvo, & B. Burge (Eds.), *Discourse, Tools and Reasoning: Essays on Situated Cognition*. (pp. 111–140). Berlin, New York: Springer.
- Goodwin, C. (1994). Professional Vision. *American Anthropologist*, 96(3), 606.
- Goodwin, C. (2003). Pointing as Situated Practice. In Sotaro Kita (ed) *Pointing: Where Language, Culture and Cognition Meet* (pp. 217–241). Mahwah NJ, Erlbaum:
- Goodwin, C. (2013). The co-operative, transformative organization of human action and knowledge. *Journal of Pragmatics*, 46(1), 8–23. <https://doi.org/10.1016/j.pragma.2012.09.003>
- Goodwin, Charles. 2018. *Co-operative Action*. Cambridge University Press.
- Hanson, N. R. (1977). *Patterns of Discovery an Inquiry Into the Conceptual Foundations of Science*. CUP Archive.
- Haviland, J. B. (2000). Pointing, Gesture Spaces, and Mental Maps. In McNeill, D. (ed.) *Language and Gesture* (pp. 13–47). Cambridge University Press.
- Heap, J. L. (2014). What Counts as Reading: Limits to Certainty in Assessment. *Curriculum Inquiry*, 10, 265–292. <https://doi.org/10.1080/03626784.1980.11075223>
- Heath, C., & Luff, P. (1992). Media Space and Communicative Asymmetries: Preliminary Observations of Video-Mediated Interaction. *Human-Computer Interaction*, 7(3), 315–346.
- Heath, C., & Luff, P. (2013). Embodied action and organisational interaction: Establishing contract on the

- strike of a hammer. *Journal of Pragmatics*, 46(1), 24–38. <https://doi.org/10.1016/j.pragma.2012.01.002>
- Henry, M. (2008). *Material Phenomenology*. Fordham Univ Press.
- Hindmarsh, J., & Heath, C. (2000). Sharing the Tools of the Trade: The Interactional Constitution of Workplace Objects. *Journal of Contemporary Ethnography*, 29(5), 523–562.
- Hindmarsh, J., Hyland, L., & Banerjee, A. (2014). Work to make simulation work: “Realism”, instructional correction and the body in training. *Discourse Studies*, 16(2), 247–269. <https://doi.org/10.1177/1461445613514670>
- Hjulstad, J. (2016). Practices of Organizing Built Space in Videoconference-Mediated Interactions. *Research on Language and Social Interaction*, 49(4), 325–341. <https://doi.org/10.1080/08351813.2016.1199087>
- Jarvis-Selinger, S., Chan, E., Payne, R., Plohman, K., & Ho, K. (2008). Clinical Telehealth Across the Disciplines: Lessons Learned. *Telemedicine and E-Health*, 14(7), 720–725. <https://doi.org/10.1089/tmj.2007.0108>
- Keevallik, L. (2013). Here in time and space: Decomposing movement in dance instruction. In *Interaction and Mobility Language and the Body in Motion*. Berlin, Boston: De Gruyter.
- Kendon, A. (2005). *Gesture: Visible Action as Utterance*. Cambridge University Press.
- Liberman, K. (2013). *More Studies in Ethnomethodology*. SUNY Press.
- Licoppe, C. (2017). Showing objects in Skype video-mediated conversations: From showing gestures to showing sequences. *Journal of Pragmatics*, 110, 63–82. <https://doi.org/10.1016/j.pragma.2017.01.007>
- Licoppe, C., Luff, P., Heath, C., Kuzuoka, H., Yamashita, N., & Tuncer, S. (2017). Showing Objects: Holding and Manipulating Artefacts in Video-mediated Collaborative Settings (pp. 5295–5306). ACM Press. <https://doi.org/10.1145/3025453.3025848>
- Licoppe, C., & Veyrier, C.-A. (2017). How to show the interpreter on screen? The normative organization of visual ecologies in multilingual courtrooms with video links. *Journal of Pragmatics*, 107, 147–164. <https://doi.org/10.1016/j.pragma.2016.09.012>
- Loenhoff, J. (2017). *Intercorporeality as a Foundational Dimension of Human Communication*. Hentet fra <http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780190210465.001.0001/acprof-9780190210465-chapter-2>
- Luff, P., Heath, C., Kuzuoka, H., Hindmarsh, J., Yamazaki, K., & Oyama, S. (2003). Fractured Ecologies: Creating Environments for Collaboration. *Human–Computer Interaction*, 18(1–2), 51–84. https://doi.org/10.1207/S15327051HCI1812_3
- Luff, P., Jirotko, M., Yamashita, N., Kuzuoka, H., Heath, C., & Eden, G. (2013). Embedded Interaction: The Accomplishment of Actions in Everyday and Video-mediated Environments. *ACM Trans. Comput.-Hum. Interact.*, 20(1), 6:1–6:22. <https://doi.org/10.1145/2442106.2442112>
- Luff, P., Heath, C., Yamashita, N., Kuzuoka, H., & Jirotko, M. (2016). Embedded Reference: Translocating Gestures in Video-Mediated Interaction. *Research on Language and Social Interaction*, 49(4), 342–361. <https://doi.org/10.1080/08351813.2016.1199088>
- Macbeth, D. (2011). Understanding understanding as an instructional matter. *Journal of Pragmatics*, 43(2), 438–451. <https://doi.org/10.1016/j.pragma.2008.12.006>
- McGuigan, N., Makinson, J., & Whiten, A. (2011). From over-imitation to super-copying: Adults imitate causally irrelevant aspects of tool use with higher fidelity than young children. *British Journal of Psychology*, 102(1), 1–18. <https://doi.org/10.1348/000712610X493115>
- Merleau-Ponty, M. (2002). *Phenomenology of Perception*. London ; New York: Routledge.
- Miller, E. A. (2003). The technical and interpersonal aspects of telemedicine: effects on doctor–patient communication. *Journal of Telemedicine and Telecare*, 9(1), 1–7. <https://doi.org/10.1258/135763303321159611>
- Miller, E. A. (2011). The continuing need to investigate the nature and content of teleconsultation communication using interaction analysis techniques. *Journal of Telemedicine and Telecare*, 17(2), 55–64. <https://doi.org/10.1258/jtt.2010.100203>
- Mondada, L. (2007). Multimodal Resources for Turn-taking: Pointing and the Emergence of Possible Next Speakers. *Discourse Studies - Quarterly-*, 9(2), 194–225.

- Mondada, L. (2011). Understanding as an Embodied, Situated and Sequential Achievement in Interaction. *Journal of Pragmatics*, 43(2), 542–552.
- Mondada, L. (2014a). Conventions for multimodal transcription. Retrieved from https://franz.unibas.ch/fileadmin/franz/user_upload/redaktion/Mondada_conv_multimodality.pdf
- Mondada, L. (2014b). Instructions in the operating room: How the surgeon directs their assistant's hands. *Discourse Studies*, 16(2), 131–161. <https://doi.org/10.1177/1461445613515325>
- Mondada, L. (2014c). The local constitution of multimodal resources for social interaction. *Journal of Pragmatics*, 65, 137–156. <https://doi.org/10.1016/j.pragma.2014.04.004>
- Nevile, M. (2015). The Embodied Turn in Research on Language and Social Interaction. *Research on Language and Social Interaction*, 48(2), 121–151. <https://doi.org/10.1080/08351813.2015.1025499>
- Nishizaka, A. (2000). Seeing What One Sees: Perception, Emotion, and Activity. *Mind, Culture, and Activity*, 7(1–2), 105–123. <https://doi.org/10.1080/10749039.2000.9677650>
- Nishizaka, A. (2014). Instructed perception in prenatal ultrasound examinations. *Discourse Studies*, 16(2), 217–246. <https://doi.org/10.1177/1461445613515354>
- Nishizaka, A. (2017). The Perceived Body and Embodied Vision in Interaction. *Mind, Culture, and Activity*, 24(2), 110–128. <https://doi.org/10.1080/10749039.2017.1296465>
- Nunberg, Geoffrey (2004). Descriptive Indexicals and Indexical Descriptions. In Marga Reimer & Anne Bezuidenhout (Eds). *Descriptions and Beyond*. Oxford: Clarendon Press, pp. 261–279.
- Peirce, C. S. (1955). *Philosophical writings of Peirce*. New York: Dover Publications.
- Raudaskoski, P. L. (1999). *The Use of Communicative Resources in Language Technology Environments: a Conversation Analytic Approach to Semiosis at Computer Media*. Hentet fra <https://vbn.aau.dk/da/publications/the-use-of-communicative-resources-in-language-technology-environ>
- Sacks, H. L., Schegloff, E. A., & Jefferson, G. (1974). A Simplest Systematics for the Organization of Turn-Taking for Conversation. *Language*, 50(4), 696–735.
- Sacks, H. og Schegloff, E. (1979). Two Preferences in the Organization of Reference to Persons in Conversation and Their Interaction. In G. Psathas (ed.): *Everyday Language: Studies in Ethnomethodology*. New York: Irvington Publishers, Inc. 15–21
- Streeck, Jürgen. (2009). Depicting gestures. *Gesture*, 9(1), 1–34.
- Streeck, J. (2010). *New Adventures in Language and Interaction*. John Benjamins Publishing.
- Streeck, Jürgen. (2017). *Self-Making Man: A Day of Action, Life, and Language*. <https://doi.org/10.1017/9781139149341>
- Stryhn, M. P., Jackson, D., & Nielsen, M. F. (2016). A pilot study of telehealth and face-to-face consultations in diagnostic audiology. *Working Papers on Interaction and Communication*, 1(3), 1–38.
- Suchman, L. A. (1987). *Plans and Situated Actions: The Problem of Human-Machine Communication*. Cambridge University Press.
- Svennevig, J. (2008). Trying the Easiest Solution First in Other-Initiation of Repair. *Journal of Pragmatics*, 40(2), 333–348.
- Svennevig, J. (2018). Decomposing Turns to Enhance Understanding by L2 Speakers. *Research on Language and Social Interaction*, 51(4), 398–416. <https://doi.org/10.1080/08351813.2018.1524575>
- Want, S. C., & Harris, P. L. (2002). How do children ape? Applying concepts from the study of non-human primates to the developmental study of 'imitation' in children. *Developmental Science*, 5(1), 1–14. <https://doi.org/10.1111/1467-7687.00194>
- Wittgenstein, L. (1953). *Philosophical investigations*. Blackwell.